

Towards Evidences of Long-Range Correlations in Seismic Activity

Douglas S. R. Ferreira*

*Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro, Paracambi, RJ, Brazil and
Geophysics Department, Observatório Nacional, Rio de Janeiro, RJ, Brazil*

Andrés R. R. Papa†

*Geophysics Department, Observatório Nacional, Rio de Janeiro, RJ, Brazil and
Instituto de Física, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brazil*

Ronaldo Menezes‡

*BioComplex Laboratory, Computer Sciences, Florida Institute of Technology, Melbourne, USA
(Dated: May 5, 2014)*

In this work, we introduce a new methodology to construct a network of epicenters that avoids problems found in well-established methodologies when they are applied to global catalogs of seisms. The new methodology involves essentially the introduction of a time window which works as a temporal filter. Our approach is more generic but for small regions the results coincide with previous findings. The network constructed with that model has small-world properties and the distribution of node connectivity follows a non-traditional function *q-Gaussian*, where scale-free properties are present. The vertices with larger connectivity correspond to the areas with the greatest earthquakes. These new results strengthen the hypothesis of long spatial and temporal correlations between earthquakes.

Among many natural disasters observed in nature, earthquakes are one of the most devastating not only by the number of lives lost but also by the economic damage they cause. The understanding of seismic phenomena is of crucial importance in engineering, and in social, geophysical, and geological sciences. Despite the vast existing knowledge about the seismic waves produced by landslides on failures, much remains to be discovered about the dynamics responsible for these events. One way to improve our understanding of seismic activity, which have their starting point in inner layers of the our planet, is the analysis of temporal series of events, from which, the probability of earthquakes occurrences may be calculated [1, 2].

Several studies have examined the phenomena under the viewpoint of complex systems, where, from nonlinear interactions between the elements of a system, complex patterns arise. In that direction, previous investigations, utilizing real data from seismic catalogs, and synthetic data from earthquake models, have analyzed spatio-temporal properties of seismicity from the perspective of non-extensive statistical mechanics [3, 4] and also using complex network theories [5], which also have been applied to better understand many biological [6, 7], social [8, 9] and technological systems [10, 11].

Abe and Suzuki [12, 13] have studied the complexity of seismic events by introducing the concept of an *earthquakes network* that they built using networks of geographical sites by taking data of successive epicenters from seismological catalogs of some active regions. Although Abe and Suzuki have named this network as *earthquakes network*, in this paper we will adopt the nomenclature *epicenters network*. We adopt a different

name for two reasons, firstly because we believe that the latter name has a better agreement with the real sense of building the network, secondly to avoid confusion with another network that we will define here, which has a completely different meaning. The epicenters network in [12, 13] was constructed by choosing a certain region of the world (e.g., California, Japan) and its respective earthquake catalog, which gives for each seismic event, the magnitude, and a set of spatial and temporal data of the hypocenter. The geographic region under consideration is then divided into small cubic cells, where a cell will become a vertex of the network if an earthquake has its epicenter therein, and two cells will be connected by a directed edge if two successive events occur in those respective cells. If two successive events occur in the same cell we have a self-edge. The network resulting from this process has been found to have non-trivial characteristics, being scale-free and small-world [14]. It is noteworthy that the same construction was made in [5, 15, 16] but by using the known model for earthquakes dynamics proposed by Olami-Feder-Christensen, which use concepts of self-organized criticality in non-conservative systems [17–19].

There is, however, an important issue concerning earthquakes that remains unsolved: the possibility to establish long-range relationships between events located spatially and temporally far apart.

Observations reported from the analysis of the great 1906 San Francisco earthquake, in California, suggest that this earthquake has induced earthquakes several hundred miles away from the rupture zone (zone of breakage) [20]. Thus, the possibility of seismic spatial long-range correlations emerges resulting in the analyses re-

stricted to small seismic areas being inappropriate or incomplete. Moreover, studies from a temporal series of earthquakes, using different ways of building the network of epicenters suggest the existence of long-range correlations between different temporal and spatial seismic events, making it inconsistent with the hypothesis of the so called aftershocks zones [21–23].

Using the same mechanism of connections for successive epicenters employed in [12, 14] for the construction of the epicenters network, but with data from the global catalog of earthquakes, Ferreira et al. [24] built a global network of epicenters, considering all earthquakes with magnitude greater than 4.5 between the years 1972 to 2011, regardless of where in the world these earthquakes have taken place. The fundamental finding was that this global network is also complex, scale-free and small-world, which can lead to the interpretation that the growth rule of this network is of the preferential attachment type as described in [25]. However, when analyzing the results obtained for the distribution of connectivity for this global network, created from the model of successive events, it was observed that the power-law behavior is maintained only for vertices with low connectivity presenting an exponential cutoff for the vertices of higher connectivity [24]. This feature indicates a loss in the connectivity (preference) of vertices with a large number of connections. Therefore, in order to corroborate the studies on seismic dynamics through the use of theories of complex networks, in this paper we present a new method for building the global network of epicenters.

First, to make possible the definition of vertices in the network to be constructed, the surface of the planet is divided in equal square cells of side $L \times L$, where a cell will become a vertex of the network every time the epicenter of an earthquake is located therein. Differently from the work of Abe and Suzuki [12, 14] and similar to our previous work [24], our cells are always $L \times L$ regardless of the position in the globe (done using Equation 1). To create the links between the vertices we use a chronologically ordered time series data of seism and then define a “time window” (w) where the vertex corresponding to the first event is connected to all vertices within this window by directed edges but respecting the time order of seisms. Thereafter, the time window is moved forward so that we restart again at the next event and the new first vertex will be connected to all vertices within this window. To illustrate, suppose that it has been adopted a time window of value equal to T in a given dataset of seismic events and that within the first window there are N_s earthquakes (s_1, s_2, \dots, s_{N_s}), where each event occurs in a cell in the globe (not necessarily distinct). Thus, we assume that there is a probable relationship between the earthquake occurred in cell c_{s_1} and the others $N_s - 1$ earthquakes within this time window, where this relationship will be represented in the network by directed edges between $c_{s_1} \rightarrow c_{s_2}, c_{s_1} \rightarrow c_{s_3}, \dots, c_{s_1} \rightarrow c_{s_{N_s-1}}$,

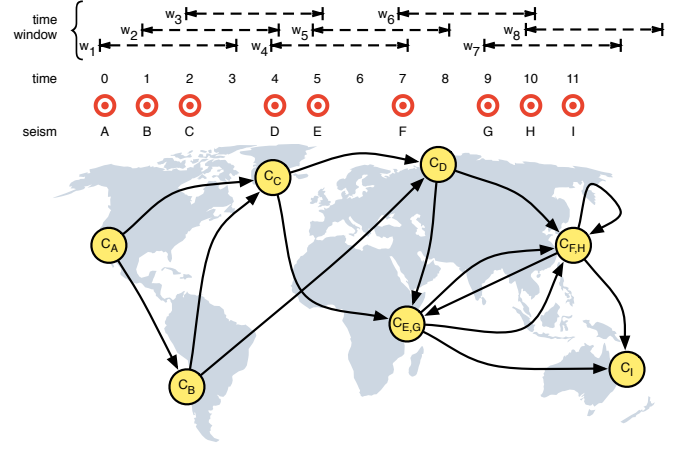


FIG. 1. Construction of the epicenter network. The time windows are represented by w_i , where i is the window number and all time windows must have the same value, in this example $T=3$ (the time is represented in arbitrary units). Events in the same window are connected as explained in the text. We can see that there are 9 earthquakes (A, B, C, D, E, F, G, H, I), but the network of epicenters has only 7 vertices ($c_A, c_B, c_C, c_D, c_E, c_F, c_I$), because $c_E = c_G$ and $c_F = c_H$. It can also be observed that $c_F \rightarrow c_H$ is a self edge.

$c_{s_1} \rightarrow c_{s_{N_s}}$. Note, however, that due to the temporal sequence we do not have a direct edge added from c_{s_k} to c_{s_j} when $k > j$; i.e. a seism s_k that happens after s_j will produce the edge $c_{s_j} \rightarrow c_{s_k}$ but not the edge $c_{s_k} \rightarrow c_{s_j}$. After all the edges are added within one time window, the window will be moved forward, starting with the cell c_{s_2} , where the earthquake s_2 occurred. Then the same procedure for adding edges described above is repeated. The construction of the network will be completed when it is no longer possible to move the window forward. It is worth noting that, as done in previous works [14, 26], we have also neglected any classification of foreshocks, main shocks and aftershocks, putting all the events on an equal footing. Figure 1 illustrates an example of the process for creating this network of epicenters. It is also noted that if two or more epicenters belonging to the same time window occur in the same cell, this cell will be connected to itself forming a self-edge.

The determination of the cell of each epicenter is made by observing the latitude ϕ_E and the longitude λ_E of each epicenter to a given reference, which, for simplicity, has been adopted as $\phi_0 = 0$ e $\lambda_0 = 0$. From these coordinates we can calculate the north-south (S_E^{ns}) and east-west (S_E^{ew}) distances between the location of the epicenter and the referential adopted. Considering the spherical approximation of the Earth we have:

$$\begin{aligned} S_E^{ns} &= R \cdot \phi_E \\ S_E^{ew} &= R \cdot \lambda_E \cdot \cos \phi_E, \end{aligned} \quad (1)$$

where $R = 6.371 \times 10^3$ km is the Earth’s radius.

After calculating these distances, we need to define the size L of the side of each square cell. It should be noted here that care must be taken when choosing the size of these cells, because if they are too small we will have a high resolution and very few seism are likely to be in the same cell leading the network to be sparse and not quite useful. On the other hand, if the cells are too large the resolution will be too small, so there will be many repetitions of events within the same cell and the network will be too dense and also not useful. Following a previous study [24], in the present paper we will use $L = 20$ km, for the sides of the square cells $L \times L$ in the globe. Thus, to obtain the location of each cell in north-south and east-west directions, we have just to divide S_E^{ns} e S_E^{ew} , respectively, by L .

The dataset used in our analyses was obtained from the World Catalogue of Earthquakes of Advanced National Seismic System (ANSS) [27] and includes seisms between 2002 and 2011. In this catalog, events whose magnitudes (m) are less than 4.5 (Richter scale) are not recorded for all parts of the world, so to obtain a more homogeneous distribution of the data for the whole world we consider only earthquakes with $m \geq 4.5$, which gives a total of 66 503 events, among which over 85% have hypocenter at depths of less than 100 km, a characteristic that allows us to use our model in the approximation of square cells, rather than cubic. The probability distribution of magnitudes of earthquakes in our data is in agreement with the Gutenberg-Richter law with a b -value exponent equal to 1.081 ± 0.004 , which was obtained by the maximum likelihood method. This result is expected, given that the Gutenberg-Richter law only presents problems for small magnitude values [3].

After the construction of the network, we performed some experiments in order to understand the structure of this network. As stated earlier, studies using the model of successive worldwide connections, show a distribution of connectivities in the form of power laws with an exponential cutoff. However, when analyzing the same data stream using the standard window of time for various values of the window, we observe two important features in the distribution of connectivities: (i) the cutoff disappears, recovering for the entire range of connectivity, the behavior type $P(k) \sim k^{-\gamma}$, leading again to the idea that the growth rule of this network follows a preferential attachment; (ii) taking into account only cells with $k > 3$, it is possible to make the distribution invariant with respect to the value of the time window by using the scale function:

$$P_{\geq}(k, T) = T^{-1} f(k/T), \quad (2)$$

where $f(x)$ decays as $x^{-\alpha}$ with $\alpha = 1.00$, which is in agreement with the scaling hypothesis. However, even when we use cells with any connectivity degrees ($k > 0$), we still have the data collapse, but in this case the best fit is obtained by a non-traditional q -Gaussian function,

$P_{\geq}(k) = B[1 - (1 - q)\beta k^2]^{1/(1-q)}$, which is a generalization of the Gaussian curve, where taking the limit $q \rightarrow 1$ we recover the standard Gaussian and for $q > 1$ it exhibits power law asymptotic behavior. It is noteworthy that the q -Gaussian distribution appears naturally from the maximization of the Tsallis entropy [28], which is used to explain many complex systems with characteristics such as long-range interaction between its elements and long-range temporal memory, where the traditional Boltzmann-Gibbs statistical mechanics does not seem to apply [29–34]. The characteristics described in (i) and (ii) are shown in Figs. 2(a), 2(b) and 2(c). In addition, we emphasize the fact that we also have calculated the γ exponent using $L = 10$ km and 15 km and the result remains unchanged with respect to the variation in cell size.

In order to show the consistency of the time window methodology, we performed three tests. The first is based on the fact that, as stated above, for small active regions on the planet, the model of successive events produce power-law distributions without cutoff [4, 12], so we used data obtained from the catalog of Southern California Earthquake Data Center (SCEDC) [35] between 2002 and 2011 to plot the graph of the distribution of connectivities for the network of epicenters of California ($32^\circ\text{N} - 37^\circ\text{N}$ and $114^\circ\text{W} - 122^\circ\text{W}$), where the total number of events was 147 435 and the cell size considered was $L = 5$ km, since $L = 20$ km is too large for a small region like California. As can be seen in Fig. 2(d), for small regions both models have similar results. The second test was to take the time series data and randomly rearrange the locations where the events occurred, while maintaining the instant they occurred. The distribution of connectivities for the network constructed from the data randomly rearranged is different from the original distribution, as shown in Fig. 2(e). In the last test we check if the value that we considered as lower threshold of magnitude is satisfactory. So we did the same analysis in the connectivity distribution using different magnitude thresholds for the globe. The magnitude intervals considered were $m \geq 4.5$, $m \geq 5.0$ and $m \geq 5.5$, comprehending 66 503, 19 038 and 5 331 earthquakes respectively [Fig. 2(f)]. The number of nodes in the epicenters network in each case are 16 682, 3 731 and 681.

To better characterize our network of epicenters it is important to study, in addition to the distribution of connectivities, other features of this network. Two important metrics in the study of complex networks are the *clustering coefficient* (C) and the *average shortest path* (ℓ). Their importance stem from the fact that they can be used to characterize small-world networks. Small-world networks are those with dense connectivity areas and long jumps between these areas. Hence, a network to be considered small-world needs to have a small *average shortest path* when compared to the number of vertices and a high *clustering coefficient* compared to a similar ran-

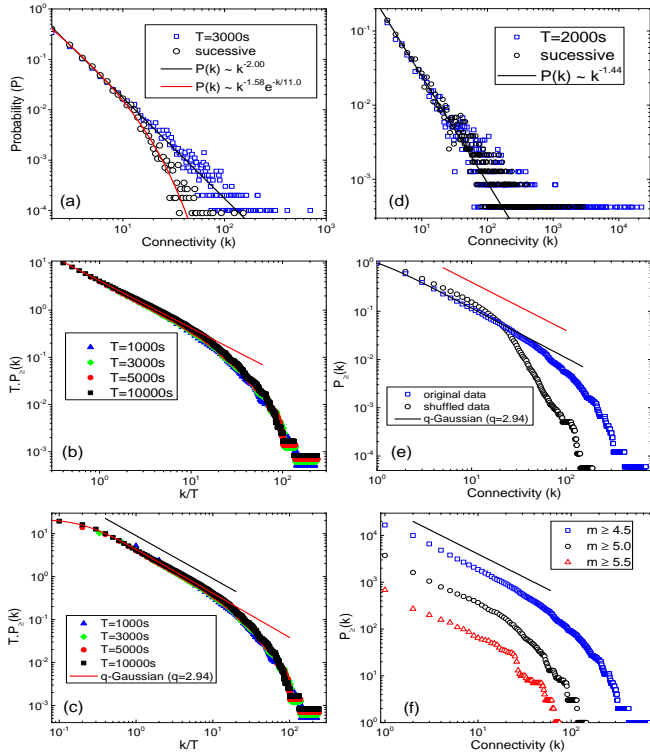


FIG. 2. Probability connectivity distribution for seismic data between 2002 and 2011. (a) Comparison between the time window model, for $T = 3000$ s, and the model of successive events for the global network with $L = 20$ km. The model of successive events obeys a power-law with exponential cutoff of the type $P(k) \sim k^{-\delta} e^{-k/k_c}$, with $\delta = 1.58$ and $k_c = 11$, while the time window model one follows a power law $P(k) \sim k^{-\gamma}$, with $\gamma = 2.00 \pm 0.01$ (calculated by the method of maximum likelihood). (b) Cumulative probability distribution for global network with $L = 20$ km and $T = 1000$ s, 3000 s, 5000 s, 10000 s, considering only epicenter with $k > 3$. The red solid line with exponent $\alpha = 1.00$ is shown as a guide. (c) The same plot as in (b), but considering epicenters with all connectivity degrees ($k > 0$). The best-fitting is for a q -Gaussian with $\beta = 11.6 \pm 0.2$ and $q = 2.94 \pm 0.01$ (red line). The black solid line has exponent $\alpha = 1.00$. (d) Comparison between the time window model, for $T = 2000$ s, and the model of successive events using the data for California with $L = 5$ km. Both models obey a power law with $\gamma = 1.44 \pm 0.02$. (e) Comparison between original data and data for a shuffled global network, both with $L = 20$ km and $T = 3000$ s. The red solid line has exponent $\alpha = 1.00$. (f) Cumulative probability distribution for earthquakes with $m \geq 4.5$, $m \geq 5.0$ and $m \geq 5.5$. The black solid line has exponent $\alpha = 1.00$. In all cases k represents in-degree.

dom network. To perform the analysis of the *clustering coefficient* and the *average shortest path* in our network, one must recall that the chosen time window will directly affect the measure of these quantities. Thus, it is natural to wonder if there is a best value for the time window. Prior to answering this question, remember that if we are trying to insert a time window in the time series data,

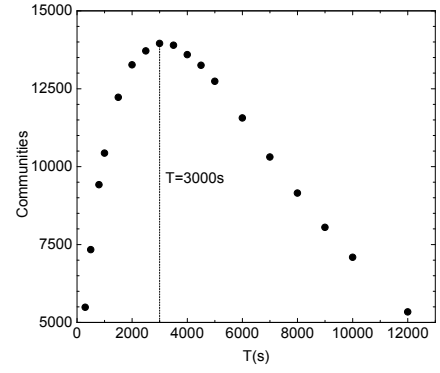


FIG. 3. Variation of the number of communities in the earthquake network with the size of the time window. The best window is found when the number of communities is maximum, i.e., at $T = 3000$ s. Seismic data from around the world between the years 2002 and 2011, for $m \geq 4.5$ were used.

this window will be inserted in the *earthquakes network*, i.e., in the network of connections between earthquakes (in the example given earlier the vertices of this network are: s_1, s_2, \dots, s_{N_s}) and not in the *epicenters network*, which is the spatial network of cells where earthquakes occurred (in the example previously given the vertices of this network are: $c_{s_1}, c_{s_2}, \dots, c_{s_{N_s}}$). We also emphasize the impossibility of loops in the earthquakes network, since an event cannot be connected to itself (different from the epicenters network, where a cell can be linked to itself, as illustrated in Fig.1). So we need to find a window not too large, that will create many links between earthquakes that probably are not correlated, but neither too small, which would bring to us a very fragmented network in which many earthquakes that have correlations between them are not connected. Thus, to find this window, we have calculated the number of communities in the networks of earthquakes and observed how this number depends on window size. In simple terms, a community is a collection of vertices in the network that is densely connected internally (inside the community). To calculate the number of communities we use the Louvain method [36]. From the results shown in Fig.3, we can confirm the fact that small windows will cause the network of earthquakes to be very fragmented and consequently with few communities, and large windows will produce very large clusters also causing a decrease in the number of communities. This leads us to the conclusion that the ideal window occurs when the number of communities is a maximum and this happens, in the case of the global earthquakes network, when $T = 3000$ s [Fig. 3], and in the case of the earthquakes network for California, when $T = 2000$ s (the plot is not presented here).

Using the ideal time window found, we built a global network of epicenters (spatial network of cells where the epicenters have occurred) using cells of size 20 km

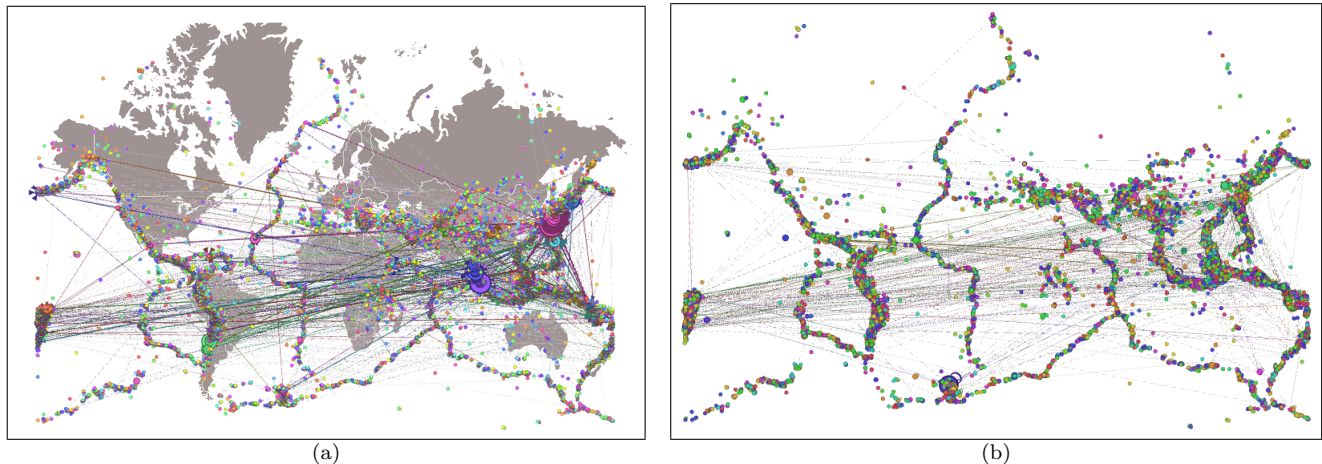


FIG. 4. Geospatial picture of the global network of cells between the years 2002 and 2011, for $L = 20$ km and $m \geq 4.5$, where it is possible to clearly observe the tectonic faults delineate by cells in the network. For clarity, only the links that occurred more than 2 times between the same two cells are shown. Larger cells have a higher number of connections and cells of the same color belong to the same community. (a) Time-window model for $T = 3000$ s. The sites with the largest cells are located around Japan, Sumatra Island and Chile. (b) The same as in (a) for the successive model.

$\times 20$ km; the number of vertices in that network is $N = 21\,158$. The measurements of C and ℓ , were performed using the algorithms described in [37] and [38], respectively. The results for our epicenters network (not to be confused with the earthquake network definition of Abe and Suzuki [12, 13]) were $C = 1.91 \times 10^{-1}$ and $\ell = 6.50$. For a random network with the same size the *clustering coefficient* has a $C_{rand} = 1.71 \times 10^{-4}$ value. From these results we observe that the global network of epicenters created from our model of time windows has small-world properties, given that $C \gg C_{rand}$ and that $\ell \ll N$ ($\ell \approx \log N$).

In our network construction it is possible to know the exact geographical location of each cell, the number of connections of each cell and to which other cells they are connected to. Therefore we used the software Gephi [39] to obtain a geo-spatial image of the seisms. In Fig. 4(a) it is possible to observe that the cells of higher connectivity are located in Japan, in the Sumatra Island and in Chile. This fact is interesting because these three regions were those with earthquakes of greater magnitudes between 2002 and 2011 [40], which makes sense since it is expected that earthquakes of great magnitudes produce more aftershocks than earthquakes of smaller magnitudes. Note however that this result is not found when using the model of successive events, where the three regions mentioned above have the same connectivity of many other places [Fig. 4(b)]; once again, we have a good indication that the time-window model is a better approach for constructing networks of epicenters.

In summary, we showed that the global network of epicenters obtained from the time-window construction has scale-free properties, where the distribution of connectiv-

ities follows a *q-Gaussian* and, if using a scaling function, it is shown to be invariant with respect to the value of time window adopted. We have also shown the consistency of the construction by three methods: first, by applying the model to a small region (California) and finding similar results to those found when using the model of successive events; second rearranging the time series data of epicenters and finding different results from those found when using original data; third, we found the same behavior for the connectivity distribution by using different magnitude thresholds. Furthermore, we have defined a mechanism for determining the best time window and observe that the network built using that window has small-world characteristics and the cells with the greatest connectivity are located in Japan, Sumatra and Chile (regions where the earthquakes of greatest magnitudes occurred between 2002 and 2011). Therewith, our results seem to contribute to the idea of a possible long-range correlation between spatially separated locations as well as a long-range temporal memory between earthquakes temporally apart from each others.

* douglas.ferreira@ifrj.edu.br

† papa@on.br

‡ rmenezes@cs.fit.edu

- [1] B. Gutenberg and C. Richter, Bull. Seism. Soc. Am. **32**, 163 (1942).
- [2] F. Omori, Coll. Sci. Imper. Univ. Tokyo **7**, 111 (1894).
- [3] A. H. Darooneh and A. Mehri, Physica A: Statistical Mechanics and its Applications **389**, 509 (2010).
- [4] S. Abe and N. Suzuki, J. Geophys. Res. **108**, 2113 (2003).

- [5] T. P. Peixoto and C. P. Prado, *Physical Review E* **69**, 025101 (2004).
- [6] I. J. G. Portillo and P. M. Gleiser, *PloS one* **4**, e6863 (2009).
- [7] A.-L. Barabási and Z. N. Oltvai, *Nature Reviews Genetics* **5**, 101 (2004).
- [8] M. E. Newman, *Proceedings of the National Academy of Sciences* **98**, 404 (2001).
- [9] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [10] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **401**, 130 (1999).
- [11] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Physical Review Letters* **87**, 258701 (2001).
- [12] S. Abe and N. Suzuki, *Europhys. Lett.* **65**, 581 (2004).
- [13] S. Abe and N. Suzuki, *Physica A* **337**, 357 (2004).
- [14] S. Abe and N. Suzuki, *Nonlin. Processes Geophys.* **13**, 145 (2006).
- [15] T. P. Peixoto and C. P. Prado, *Physica A: Statistical Mechanics and its Applications* **342**, 171 (2004).
- [16] T. P. Peixoto and C. P. Prado, *Physical Review E* **74**, 016126 (2006).
- [17] Z. Olami, H. J. S. Feder, and K. Christensen, *Physical Review Letters* **68**, 1244 (1992).
- [18] K. Christensen and Z. Olami, *Physical Review A* **46**, 1829 (1992).
- [19] Z. Olami and K. Christensen, *Physical Review A* **46**, R1720 (1992).
- [20] D. W. Steeples and D. Steeples, *Bull. Seismol. Soc. Am.* **86**, 921 (1996).
- [21] M. Baiesi and M. Paczuski, *Phys. Rev. E* **69** (2004).
- [22] M. Baiesi and M. Paczuski, *Nonlin. Processes Geophys.* **12**, 1 (2005).
- [23] S. Abe and N. Suzuki, *Eurphys. Lett* **97** (2012).
- [24] D. Ferreira and R. Papa, A.R.R. and Menezes, *Physica A: Statistical Mechanics and its Applications* - DOI: 10.1016/j.physa.2014.04.024 (2014).
- [25] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [26] J. Davidsen and M. Paczuski, *Phys. Rev. Lett.* **94** (2005).
- [27] [Http://quake.geo.berkeley.edu/anss](http://quake.geo.berkeley.edu/anss).
- [28] C. Tsallis, *Journal of Statistical Physics* **52**, 479 (1988).
- [29] F. Vallianatos and P. Sammonds, *Tectonophysics* **590**, 52 (2013).
- [30] C. S. Barbosa, D. S. Ferreira, M. A. do Espírito Santo, and A. R. Papa, *Physica A: Statistical Mechanics and its Applications* **392**, 6554 (2013).
- [31] J. Ludescher, C. Tsallis, and A. Bunde, *EPL (Europhysics Letters)* **95**, 68002 (2011).
- [32] G. Livadiotis, D. McComas, M. Dayeh, H. Funsten, and N. Schwadron, *The Astrophysical Journal* **734**, 1 (2011).
- [33] G. Aad, B. Abbott, J. Abdallah, A. Abdelalim, A. Abdesselam, O. Abidinov, B. Abi, M. Abolins, H. Abramowicz, H. Abreu, et al., *New Journal of Physics* **13**, 053033 (2011).
- [34] F. Tamarit, S. Cannas, and C. Tsallis, *The European Physical Journal B-Condensed Matter and Complex Systems* **1**, 545 (1998).
- [35] [Http://www.data.scec.org/](http://www.data.scec.org/).
- [36] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *J. Stat. Mech. Theor. Exp.* **10**, 1000 (2008).
- [37] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3747 (2004).
- [38] U. Brandes, *Journal of Mathematical Sociology* **25**, 163 (2001).
- [39] [Https://gephi.org/](https://gephi.org/).
- [40] [Https://earthquake.usgs.gov/earthquakes/eqarchives/](https://earthquake.usgs.gov/earthquakes/eqarchives/).